

RECEIVED

APR 22

TECH CENTER 1605/2500

OIPE

RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/622,419

TIME: 13:27:52

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

C-->

ENTERED

1 <110> APPLICANT: Schroeder, Hartwig
 3 <120> TITLE OF INVENTION: Process for preparing biotin
 5 <130> FILE REFERENCE: 0000048792/UP
 7 <140> CURRENT APPLICATION NUMBER: US 09/622,419
 8 <141> CURRENT FILING DATE: 2003-03-13
 10 <150> PRIOR APPLICATION NUMBER: PCT/EP 99/01052
 11 <151> PRIOR FILING DATE: 1999-02-17
 13 <160> NUMBER OF SEQ ID NOS: 17
 15 <170> SOFTWARE: PatentIn Release 1.0, version 1.25 (EPO)/WordPerfect
 17 <210> SEQ ID NO: 1
 18 <211> LENGTH: 1155
 19 <212> TYPE: DNA
 20 <213> ORGANISM: Escherichia coli
 22 <220> FEATURE:
 23 <221> NAME/KEY: CDS
 24 <222> LOCATION: 1..1155
 26 <400> SEQUENCE: 1
 28 atg gca aaa cac ctt ttt acg tcc gag tcc gtc tct gaa ggg cat cct 48
 29 Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
 30 1 5 10 15
 32 gac aaa att gct gac caa att tct gat gcc gtt tta gac gcg atc ctc 96
 33 Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
 34 20 25 30
 36 gaa cag gat ccg aaa gca cgc gtt gct tgc gaa acc tac gta aaa acc 144
 37 Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
 38 35 40 45
 40 ggc atg gtt tta gtt ggc ggc gaa atc acc acc agc gcc tgg gta gac 192
 41 Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
 42 50 55 60
 44 atc gaa gag atc acc cgt aac acc gtt cgc gaa att ggc tat gtg cat 240
 45 Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
 46 65 70 75 80
 48 tcc gac atg ggc ttt gac gct aac tcc tgt gcg gtt ctg agc gct atc 288
 49 Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
 50 85 90 95
 52 ggc aaa cag tct cct gac atc aac cag ggc gtt gac cgt gcc gat ccg 336
 53 Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
 54 100 105 110
 56 ctg gaa cag ggc gcg ggt gac cag ggt ctg atg ttt ggc tac gca act 384
 57 Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
 58 115 120 125
 60 aat gaa acc gac gtg ctg atg cca gca cct atc acc tat gca cac cgt 432
 61 Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg

RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/622,419

TIME: 13:27:52

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

62	130	135	140	
64	ctg gta cag cgt cag gct gaa gtg cgt aaa aac ggc act ctg ccg tgg	480		
65	Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp			
66	145 150 155 160			
68	ctg cgc ccg gac gcg aaa agc cag gtg act ttt cag tat gac gac ggc	528		
69	Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly			
70	165 170 175			
72	aaa atc gtt ggt atc gat gct gtc gtg ctt tcc act cag cac tct gaa	576		
73	Lys Ile Val Gly Ile Asp Ala Val Leu Ser Thr Gln His Ser Glu			
74	180 185 190			
76	gag atc gac cag aaa tcg ctg caa gaa gcg gta atg gaa gag atc atc	624		
77	Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile			
78	195 200 205			
80	aag cca att ctg ccc gct gaa tgg ctg act tct gcc acc aaa ttc ttc	672		
81	Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe			
82	210 215 220			
84	atc aac ccg acc ggt cgt ttc gtt atc ggt ggc cca atg ggt gac tgc	720		
85	Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys			
86	225 230 235 240			
88	ggt ctg act ggt cgt aaa att atc gtt gat acc tac ggc ggc atg gcg	768		
89	Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala			
90	245 250 255			
92	cgt cac ggt ggc ggt gca ttc tct ggt aaa gat cca tca aaa gtg gac	816		
93	Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp			
94	260 265 270			
96	cgt tcc gca gcc tac gca gca cgt tat gtc gcg aaa aac atc gtt gct	864		
97	Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala			
98	275 280 285			
100	gct ggc ctg gcc gat cgt tgt gaa att cag gtt tcc tac gca atc ggc	912		
101	Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly			
102	290 295 300			
104	gtg gct gaa ccg acc tcc atc atg gta gaa act ttc ggt act gag aaa	960		
105	Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys			
106	305 310 315 320			
108	gtg cct tct gaa caa ctg acc ctg ctg gta cgt gag ttc ttc gac ctg	1008		
109	Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu			
110	325 330 335			
112	cgc cca tac ggt ctg att cag atg ctg gat ctg ctg cac ccg atc tac	1056		
113	Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr			
114	340 345 350			
116	aaa gaa acc gca gca tac ggt cac ttt ggt cgt gaa cat ttc ccg tgg	1104		
117	Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp			
118	355 360 365			
120	gaa aaa acc gac aaa gcg cag ctg ctg cgc gat gct gcc ggt ctg aag	1152		
121	Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys			
122	370 375 380			
124	taa	1155		
127	<210> SEQ ID NO: 2			
128	<211> LENGTH: 384			

RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/622,419

TIME: 13:27:52

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

```

129 <212> TYPE: PRT
130 <213> ORGANISM: Escherichia coli
132 <400> SEQUENCE: 2
135 Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
136   1           5           10           15
138 Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
139           20           25           30
141 Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
142           35           40           45
144 Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
145   50           55           60
147 Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
148   65           70           75           80
150 Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
151           85           90           95
153 Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
154           100          105          110
156 Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
157           115          120          125
159 Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
160           130          135          140
162 Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
163 145          150          155          160
165 Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
166           165          170          175
168 Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
169           180          185          190
171 Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
172           195          200          205
174 Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
175           210          215          220
177 Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
178 225          230          235          240
180 Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
181           245          250          255
183 Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
184           260          265          270
186 Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
187           275          280          285
189 Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
190           290          295          300
192 Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
193 305          310          315          320
195 Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
196           325          330          335
198 Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
199           340          345          350
201 Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
202           355          360          365

```

RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/622,419

TIME: 13:27:52

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

```

204 Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
205      370                      375                      380
209 <210> SEQ ID NO: 3
210 <211> LENGTH: 1206
211 <212> TYPE: DNA
212 <213> ORGANISM: Escherichia coli
214 <220> FEATURE:
215 <221> NAME/KEY: CDS
216 <222> LOCATION: 1..1206
218 <400> SEQUENCE: 3
221 atg aac gtt ttt aat ccc gcg cag ttt cgc gcc cag ttt ccc gca cta      48
222 Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
223   1          5          10          15
225 cag gat gcg ggc gtc tat ctc gac agc gcc gcg acc gcg ctt aaa cct      96
226 Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
227          20          25          30
229 gaa gcc gtg gtt gaa gcc acc caa cag ttt tac agt ctg agc gcc gga      144
230 Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
231          35          40          45
233 aac gtc cat cgc agc cag ttt gcc gaa gcc caa cgc ctg acc gcg cgt      192
234 Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
235          50          55          60
237 tat gaa gct gca cga gag aaa gtg gcg caa tta ctg aat gca ccg gat      240
238 Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
239   65          70          75          80
241 gat aaa act atc gtc tgg acg cgc ggc acc act gaa tcc atc aac atg      288
242 Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
243          85          90          95
245 gtg gca caa tgc tat gcg cgt ccg cgt ctg caa ccg ggc gat gag att      336
246 Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
247          100          105          110
249 att gtc agc gtg gca gaa cac cac gcc aac ctc gtc ccc tgg ctg atg      384
250 Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
251          115          120          125
253 gtc gcc caa caa act gga gcc aaa gtg gtg aaa ttg ccg ctt aat gcg      432
254 Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
255          130          135          140
257 cag cga ctg ccg gat gtc gat ttg ttg cca gaa ctg att act ccc cgt      480
258 Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
259 145          150          155          160
261 agt cgg att ctg gcg ttg ggt cag atg tcg aac qtt act ggc ggt tgc      528
262 Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
263          165          170          175
265 ccg gat ctg gcg cga gcg att acc ttt gct cat tca gcc ggg atg gtg      576
266 Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
267          180          185          190
269 gtg atg gtt gat ggt gct cag ggg gca gtg cat ttc ccc gcg gat gtt      624
270 Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
271          195          200          205

```

RAW SEQUENCE LISTING

DATE: 03/19/2003

PATENT APPLICATION: US/09/622,419

TIME: 13:27:52

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

```

273 cag caa ctg gat att gat ttc tat gct ttt tca ggt cac aaa ctg tat      672
274 Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
275      210      215      220
277 ggc ccg aca ggt atc ggc gtg ctg tat ggt aaa tca gaa ctg ctg gag      720
278 Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
279 225      230      235      240
281 gcg atg tcg ccc tgg ctg ggc ggc ggc aaa atg gtt cac gaa gtg agt      768
282 Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
283      245      250      255
285 ttt gac ggc ttc acg act caa tct gcg ccg tgg aaa ctg gaa gct gga      816
286 Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
287      260      265      270
289 acg cca aat gtc gct ggt gtc ata gga tta agc gcg gcg ctg gaa tgg      864
290 Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
291      275      280      285
293 ctg gca gat tac gat atc aac cag gcc gaa agc tgg agc cgt agc tta      912
294 Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
295      290      295      300
297 gca acg ctg gcg gaa gat gcg ctg gcg aaa cgt ccc ggc ttt cgt tca      960
298 Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
299 305      310      315      320
301 ttc cgc tgc cag gat tcc agc ctg ctg gcc ttt gat ttt gct ggc gtt      1008
302 Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
303      325      330      335
305 cat cat agc gat atg gtg acg ctg ctg gcg gag tac ggt att gcc ctg      1056
306 His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
307      340      345      350
309 cgg gcc ggg cag cat tgc gct cag ccg cta ctg gca gaa tta ggc gta      1104
310 Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
311      355      360      365
313 acc ggc aca ctg cgc gcc tct ttt gcg cca tat aat aca aag agt gat      1152
314 Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
315      370      375      380
317 gtg gat gcg ctg gtg aat gcc gtt gac cgc gcg ctg gaa tta ttg gtg      1200
318 Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
319 385      390      395      400
321 gat taa      1206
322 Asp
326 <210> SEQ ID NO: 4
327 <211> LENGTH: 401
328 <212> TYPE: PRT
329 <213> ORGANISM: Escherichia coli
331 <400> SEQUENCE: 4
333 Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
334 1      5      10      15
336 Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
337      20      25      30
339 Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
340      35      40      45

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/622,419

DATE: 03/19/2003

TIME: 13:27:53

Input Set : A:\622419sq

Output Set: N:\CRF4\03192003\I622419.raw

L:8 M:271 C: Current Filing Date differs, Replaced Current Filing Date